

Interference of speech and interior noise of Chinese high-speed trains with task performance

Weigang Wei^{a,*}, Annelies Bockstael^a, Bert De Coensel^a, Dick Botteldooren^a

*^aGhent University, Department of Information Technology, Acoustics research group,
St.-Pietersnieuwstraat 41, B-9000 Ghent, Belgium*

Abstract

High speed trains are promoted all over the world as an alternative to aircraft (or conventional trains) for travel distances from a few hundred to a (few) thousand kilometers. Long travel time broadens the scale of activities potentially affected by train noise and sounds from other passengers. In this experiment, the effect on task performance caused by the combination of train noises and passengers' chatting was investigated. While being exposed to different noise combinations, Chinese as well as Dutch speaking participants were asked to perform a series of calculation and memory tasks as quickly and as accurately as possible. Results mainly show that (i) response time and accuracy are independent variables for both tasks, with the former showing the largest effect caused by background noise; (ii) a model based on saliency and attention can predict a significant part of the influence of combined noise (train and speech) on task performance; and (iii) in retrospect, the majority of participants report the speech sound to have been either equally or more annoying although the addition of speech changed the overall L_{Aeq} by less than 0.1 dB(A).

*Corresponding author. Tel.: +32 9 264 9995; fax: +32 9 264 9969.
Email address: weigang@intec.ugent.be (Weigang Wei)

Keywords: Train noise, Noise annoyance , Task performance, Attention model

1. Introduction

The development of modern high-speed trains started in the 1960's, with the construction of the Japanese Tōkaidō Shinkansen line between Tokyo and Osaka, on which traveling speeds of 210 km/h were reached. Because of growing environmental interest and rising road traffic congestion, European countries soon followed, with the development of high-speed train lines in France (TGV), Germany and Italy. The prominent advantages of high traveling speeds and extensive goods carrying capabilities, increasingly stimulate policy makers and industry to invest in the development of new lines [1, 2, 3]. For example, to improve public transport capabilities, China started to invest in high-speed train projects from the beginning of the 21st century. In 2009, the construction of the Wuhan-Guangzhou Line was completed, connecting more than twenty cities and 100 million people.

Despite their advantages, high-speed trains often bring more noise problems to both passengers and residents along the rail lines than conventional trains. Researchers have investigated the acoustical characteristics of Japanese [4, 5, 6] and European high-speed trains [7, 8, 9, 10]. Additionally, annoyance caused by the noise of high-speed trains has been studied extensively [11, 12, 13, 14]. To date, few data about the noise and annoyance caused by Chinese high-speed trains have been published [15]. Although the mechanical systems of Chinese high-speed trains are similar to those in Europe, differences in the manufacturing process and the finish of vehicles may result in significant differences in internal and external noise levels and spectra. Due to cultural differences and the presence of mountainous terrain, the amount of noise annoyance caused by Chinese high speed railway

lines to residents along the track may also differ from that of Europe. Another issue is that distances and traveling times in China are much longer than in Europe, giving rise to a wider variety of social activities occurring inside the trains. A survey shows that about 64% of passengers in Chinese high-speed trains engage in reading activities [16]. Furthermore, the availability of electrical facilities in the new Chinese trains was found to lead to passengers using their personal computer to work during traveling [17]. Therefore, when assessing the noise environment inside Chinese high-speed train cabins, the combination of train noise and other sound sources such as people talking is of particular interest.

It is well-known that exposure to background noise might have a negative impact on task performance [18]. In this, speech and other time varying sound are believed to have obligatory access to memory, meaning that they are recorded and processed even when attention is directed elsewhere. However, literature review [19] clearly shows that the (mechanism of) speech influence depends on the task to be performed: short-term memory tasks appear more sensitive to the phonological aspects of speech whereas reading comprehension is more influenced by the semantic content and distraction. This distinction between phonology and semantics might also explain why non-understandable speech (non-words, foreign language) has sometimes less influence than irrelevant speech spoken in the listener's mother tongue, whereas other experiments find no difference. Finally, it seems that the detrimental effect of irrelevant speech on performance can be reduced when people have time to habituate to it, but the orientation reaction, triggered by speech after a silent phase, cannot be eliminated.

In relation to speech communication and noise annoyance in Shinkansen train cabins, an optimum level of compartment noise between 50 and 60dB(A) was

recommended [20]. At this level, the train noise could partly mask the interfering speech noise and thus mitigate the feeling of annoyance. For higher sound pressure levels of 63dB(A), a slightly negative effect on annoyance was found [21, 22]. Nevertheless, measurements show that the sound pressure level inside Chinese high-speed train cabins varies between 75 and 80dB(A) in most conditions [15]. The characteristics of the background train noise vary only slightly with a change in position. However, the position of source and receiver is more critical for interfering speech, as the sound intensity of the interfering speech decreases with increasing distance. Longer distance can reduce auditory distraction by smoothing the perceived variation of sound characters [23, 24, 25] and by decreasing the signal-to-noise ratio [26, 27]. Auditory attention is influenced by other factors such as reverberation [28], the familiarity with the languages of the interfering speech [27, 29, 30] and combined stimuli. Moreover, substantial inter-subject variation exists [31, 32, 33, 34]. Additionally, noise events leading to auditory distraction have a strong relation with annoyance [35, 36, 37]. Accordingly, testing the effect of different receiver positions could give some insight.

Taking these factors in account, an experiment was carried out that investigates the effects of the sound environment inside Chinese high-speed trains on task performance, emotional state and noise annoyance of passengers. The aim of this experiment is to find out how these factors are affected by compartment noise (i) originating from the driving high-speed train only and (ii) originating from the combination of the driving high-speed train and the chatting of other passengers. Four experimental conditions are considered: (i) a baseline low noise level condition, which will be used to remove the influence of inter-individual differences, (ii) exposure to compartment noise originating from the driving high-speed train

only, (iii) exposure to mixed noise originating from the driving high-speed train and speech of foreign (non-understandable) language, and (iv) exposure to mixed noise originating from the driving high-speed train and speech of native language.

To examine the effect of noise on task performance, a number of tasks, such as prose recall, serial recall or mental arithmetics are commonly used [24, 38, 39]. Because of the sensitivity of the memory component to irrelevant sound, serial recall tasks are generally preferred [40, 41]. The studies by Banbury and Berry [39] show that meaningful noise can negatively affect both prose memory and mental arithmetic tasks, whereas meaningless noise does not affect memory task performance. In the present experiment, a calculation and a memory recall task are chosen, since these tasks do not require training. Although a change in response accuracy, as measured by the correct answer rate, might be a good indicator for a potential effect of noise on task performance [42], it might blur more subtle effects, because the background noise might not affect the correct answer rate as such, but cause the participants to require longer thinking. Therefore, the present experiment allows the subjects to spend as much time as they require on a task in order to obtain a higher correct answer rate, and thus, response times and individual exposures are also recorded.

2. Methodology

2.1. Participants

In total, 83 subjects (37 females, 46 males) participated in the experiment. Out of these, 46 had Chinese as their mother tongue, and 37 Dutch; all of them lived in Belgium during the time when the experiment was conducted. The average age of the participants was 25.0 year (standard deviation = 4.3). Because

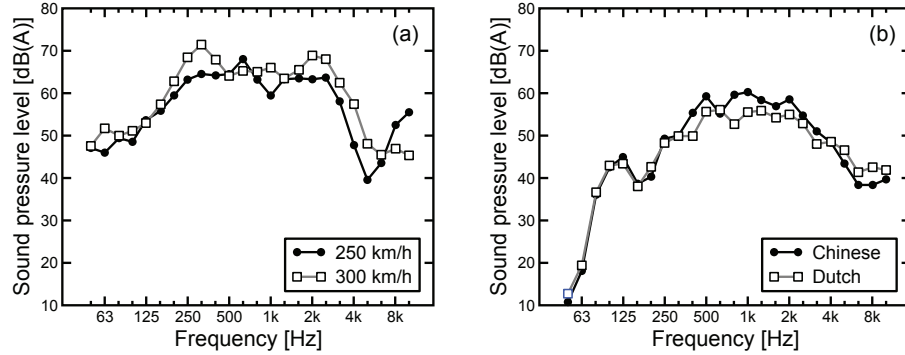


Figure 1: Spectra in 1/3-octave bands of (a) the train noise and (b) the speech sound excerpts.

participants were recruited by posting advertisements on the university's bulletin boards, all of them were students (26 bachelor, 57 postgraduate). All participants reported normal hearing ability. Additionally, participants that are fluent in Dutch as well as Chinese were not considered. After completion of the listening test, the participants received a cinema ticket as compensation.

2.2. Stimuli

The stimuli of the experiment were constructed by combining the noise inside a compartment of a traveling Chinese high-speed train with various speech fragments. Realistic binaural signals of train noise were recorded inside a high-speed train on the Wuhan-Guangzhou Line, traveling at speeds of 250 and 330 km/h, using a B&K head and torso simulator (HATS) type 4100, together with the B&K Pulse system. The compartment was empty apart from standard cabin furniture and the measurement equipment. Two representative 10-minute excerpts were selected, one for each traveling speed; Figure 1(a) shows the spectrum of both excerpts.

Binaural recordings of speech were constructed in two steps. First, the speech of people chatting was recorded in an anechoic room using a central free-field microphone. Speech utterances consisted of transcribed train conversations, originally in Chinese and translated to Dutch. The conversations were held by four nonprofessional persons having respectively Dutch or Chinese as mother tongue. Two representative 10-minute excerpts were selected, one for each language; spectra are shown in Figure 1(b). Second, the binaural impulse response in a typical Chinese high-speed train compartment was simulated using the Odeon room acoustics software [43]; Figure 2 shows the simulated cabin model. In this model, an internal source of Odeon, “Tlknorm_NATURAL”, was used to simulate a human’s voice with total sound power level 67.4 dB(A) [43, 44]. This source was put in the center of the compartment. According to the simulation, the source power level would decrease to 58.5 dB(A) in the first row, 54.8 dB(A) in the third row and 50.3 dB(A) in the sixth row. A pre-test showed that in the first row the participants focussed strongly on the speech that was clearly understandable and dominant. Since our interest is mainly in interaction between the sounds, the first listener location (labeled “near”) was placed a little further than the speaker, three rows in front of the source and the second listener location (labeled “far”) was placed six rows away from the source. Binaural impulse responses of the compartment were calculated for both configurations, and convolved with the two speech excerpts, resulting in four binaural speech recordings, labeled “nearDut”, “farDut”, “nearChn” and “farChn” to distinguish between the position of the listener and the language of the conversation (Dutch or Chinese).

Four conditions for mixing the stimuli were used, namely: the baseline condition (“CBaseline”) without train noise, the condition in which only train noise is

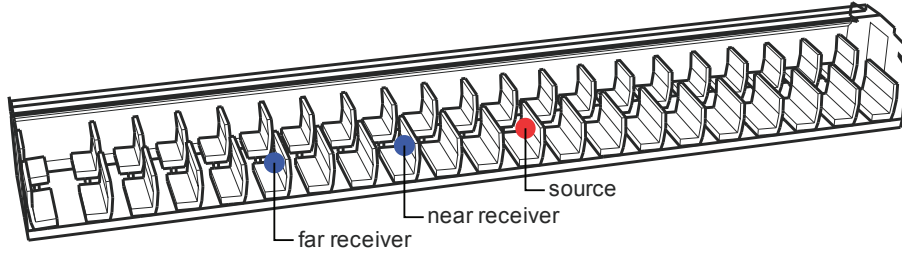


Figure 2: Simulation model of the high-speed train compartment, showing the positions of the source and two receivers.

present (“CTrain”), and the conditions in which train noise is mixed with Dutch speech (“CDutch”) and with Chinese speech (“CChinese”). Subsequently, a series of nine 10-minute binaural stimuli were constructed; an overview is given in Table 1. The first stimulus, labeled “SBaseline”, represents the baseline condition, consisting of only white noise at low sound pressure level. The stimuli “S250train” and “S330train” only consist of the high-speed train noise, without interfering speech. Additionally, six stimuli were created by combining train noise (at 250 or 330 km/h) with the various speech recordings (near or far position, Chinese or Dutch spoken). For example, the stimulus “S250nearDut” consists of train noise at 250 km/h mixed with the Dutch speech fragment, spoken at three rows distance from the listener. Finally, 3 menus of 4 stimuli each were constructed, labeled “M250near”, “M250far” and “M330near”. Each menu consists of a single stimulus from each condition, having the same train speed and source/receiver distance. Table 1 lists the main acoustical properties of the different stimuli, together with the composition of the menus.

Table 1: Composition of the experimental menus, together with the main acoustical parameters of the experimental stimuli (energetic averages between left and right ear).

Condition	Stimulus	$L_{Aeq,10min}$	L_{A10}	Menu composition		
		[dB(A)]	[dB(A)]	M250near	M250far	M330near
CBaseline	SBaseline	30.0	30.0	•	•	•
CTrain	S250train	74.6	76.0	•	•	
	S330train	77.7	79.4			•
CDutch	S250nearDut	74.6	76.5	•		
	S250farDut	74.6	76.1		•	
	S330nearDut	77.7	79.5			•
CChinese	S250nearChn	74.6	76.5	•		
	S250farChn	74.6	76.2		•	
	S330nearChn	77.7	79.5			•

2.3. Apparatus

During the experiment, subjects were seated in a sound attenuated chamber. Auditory stimuli were presented through a closed-type headphone, and the task presentation and response collection was controlled using Matlab with graphical user interface. The playback equipment was calibrated beforehand using the HATS system. All instruction and tests were conducted in the mother tongue of the subjects. The participants' responses were monitored using an additional screen, placed in an adjacent room, in order to ensure the validity of the test.

2.4. Procedure

Figure 3 shows, from left to right, the chronology of the experimental procedure. Before the start of the experiment, participants were told to complete the tests as quickly and as accurately as possible, without revealing the purpose of the noise they would hear. They were also informed that the experiment was going to be monitored by video, in order to prevent cheating. The experiment itself started with a short training session, demonstrating the participants how to perform both calculation and memory tasks (see Section 2.5). Each task consisted of four parts, and took about 20 minutes to complete. Immediately after each task, the participants had to rate how difficult they found the task, using a 5-point scale, ranging from very easy to very difficult. After both calculation and memory tasks were done (the order of the tasks was randomized among participants), a small questionnaire was administered to the participants, containing questions on age, gender, educational background, noise sensitivity and perceived noise annoyance during the test (on a 5-point scale).

Participants were randomly assigned one of the three menus, and were presented this menu twice, once for each task. Within each task, a different, ran-

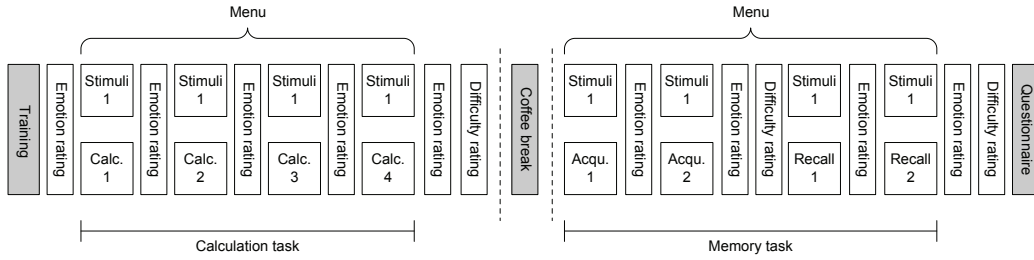


Figure 3: Experimental design. The order of the calculation task and the memory task was randomized among participants.

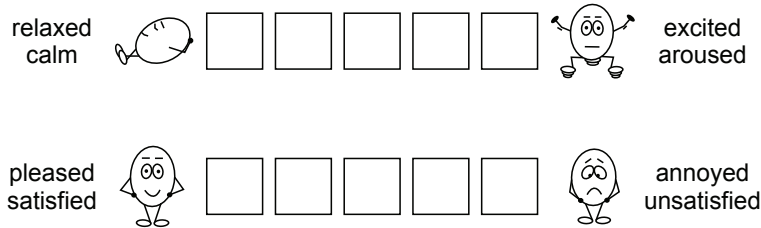


Figure 4: Emotion manikins used in the experiment, to indicate the emotion state (adapted from [46]).

domly selected stimulus from the assigned menu was presented during each of the calculation, acquisition and recall parts, see figure 3. Note that participants were only exposed to the stimuli for the time period they needed to perform the instructed part of the task, which was always less than 10 minutes. Before the start of the experiment, and after each part of each task, participants were asked to describe their emotional state using a semantic differential combined with emotion manikins [45, 46]. Two semantic differentials were used to characterize the emotional state: relaxed-excited (first row of Figure 4), and pleased-annoyed (second row of Figure 4).

2.5. Tasks

For this experiment, a calculation task (mental arithmetic) as well as a memory task (remembering sentences) are selected. Each part of the calculation task consists of 12 sums [39]. For each sum, 11 randomly generated, single-digit numbers are presented sequentially, and at the end, the participant has to enter the sum of all numbers. The participant is not allowed to take notes or count aloud. One digit is presented per mouse click, and exact clicking times are registered, as well as the final result.

The memory task consists of 2 acquisition parts and 2 recall parts, entirely carried out in the participant’s mother tongue. During the acquisition parts, 24 sentences are presented sequentially, and the content of the sentences has to be remembered. Only everyday sentences were chosen without any cultural dependent topics. The presentation of each sentence follows a particular pattern: (i) the sentence is shown; (ii) after clicking, a screen with four different colors is shown, in order to distract the participant [27]; (iii) a small question about the sentence is asked, and the participant has to fill in the answer. For example, a sentence could be “The computer was bought at the shopping center”, and a related small question could be “Where?”. Four types of sentences are used in the experiment, labeled “ones”, “twos”, “threes” and “fours” [38, 27], depending on the amount of information in the sentence. An example of a “fours” sentence could be “A white mouse ate one piece of the cheesecake on the table”, as this sentence contains four pieces of information: color (white), how much (one piece), what (cheesecake) and where (on the table). If one piece of information is removed, we get a “threes” sentence, e.g. “A mouse ate one piece of the cheesecake on the table”. By removing more information, one gets “twos” (e.g. “A mouse ate one piece

of the cheesecake”) and “ones” (e.g. “A mouse ate the cheesecake”) sentences. Similar sentences can be generated by removing information in different order. Furthermore, a “fours” sentence can be divided into different “ones”, “twos” or “threes”. Finally, the same sentences are presented to all participants in the same order, and successive sentences do not come from the same “fours” sentence.

During the recall parts, the 24 presented sentences are mixed with 12 new ones, and these are shown one by one. The participants are asked if this sentence was seen before, and how confident they feel about the answer (assessed using a 5-point scale). Additionally, exact clicking times are registered.

For analyzing the results, proper statistical methods are necessary. To compare the relation between the response and the noise, Pearson correlation coefficient will be used and to compare the difference between two parameters, t-test will be used. For the t-test, the null hypothesis states that the participants do not respond significantly different when the considered parameter varies.

3. Results

To investigate the relation between task performance and noise exposure, several steps are followed in this section. First, the relation between correct answer rate and response time is investigated. Then, the baseline is used as a reference to account for subject-related factors. Subsequently, the relation between task performance and stimuli and acoustical parameters is considered. Finally, annoyance results are analyzed.

3.1. Correlation between response accuracy and duration

Table 2 shows the Pearson’s correlation between response duration and accuracy on elementary calculations or sentence recalls (accuracy is coded 1 for a cor-

Table 2: Pearson’s correlation (and associated p -value) between response duration and accuracy, considering the different experimental conditions and task types separately.

Condition	Task	
	Calculation	Memory (recall)
CBaseline	-0.018 ($p = 0.871$)	0.011 ($p = 0.947$)
CTrain	0.036 ($p = 0.748$)	0.173 ($p = 0.432$)
CDutch	0.058 ($p = 0.602$)	0.053 ($p = 0.741$)
CChinese	0.128 ($p = 0.247$)	-0.039 ($p = 0.795$)

rect answer and 0 otherwise), for all participants, considering the different experimental conditions and task types separately. The results suggest that the correct answer rate and the corresponding response duration are independent from each other, with very low and non-significant correlation ($\alpha = 0.05$). This supports the idea that both response duration and accuracy can be analyzed separately; both variables will therefore be used as indicators for noise effects in this paper.

3.2. Baseline condition effects

Personal factors such as mental arithmetic abilities, age or gender could affect task performance in general. Therefore, a baseline condition was included in the experiment. A strong correlation between task performance in the baseline condition and in the other conditions was found. For example, for duration of the calculation task, Pearson’s correlation coefficients of 0.91, 0.93 and 0.86 ($p < 0.001$ in each case) are found between the CBaseline condition and the CTrain, CDutch and CChinese conditions respectively. Consequently, the effect

of the baseline condition was removed from the results before further analysis.

3.3. Relation between stimulus parameters and task performance

First, the effect of basic stimulus parameters (train speed, language of speech, and distance between receiver and interfering speech source) is considered. The influence of train speed is investigated by comparing response times (normalized by baseline) between the cases in which participants are exposed to the S250train and S330train stimuli. For the calculation task, an independent samples student t -test showed that the participants do not respond significantly slower or faster when train speed is varied. For the recall parts of the memory task (in which words had to be typed in), response times were found to be affected by the typing speed of the participants, even after accounting for the baseline condition, and are therefore not considered in this work.

The combined effect of language of speech and distance between interfering speech source and the listener is analyzed by comparing response times for elementary sums (the time between two mouse clicks) between the different menus. Results were averaged over all test parts performed by participants with the same mother tongue and are shown in table 3. For illustration, results are listed for 50% and 100% completion of the test. The first three rows of Table 3 show results for a train speed of 250 km/h and with the listener in the far position. It can be seen that, in this case, presenting speech facilitates the task: both the Dutch-speaking and the Chinese-speaking participants show a reduced answering time when the background noise contains language elements. Results are analyzed statistically using a pairwise t -test. It was found that the differences in answering times are only significant ($p < 0.05$) for the Chinese-speaking group. No significant effect of speech language is found. The middle three rows of Table 3 show results for

the listener in the near position. In this case, presenting speech reduces the calculation speed, but effects are only significant ($p < 0.05$) when the language of the interfering speech equals the participants mother tongue. These results can be understood by considering the signal-to-noise ratio (SNR) between the speech and the train sound: for the listener in the near position, the SNR will be higher, and therefore, one would expect that the interfering effect of the speech would also be higher in this situation. Finally, the last three rows of Table 3 show the results for a train speed of 330 km/h. On the one hand, it can be seen that, in this case, the effect of the speech disappears, because the higher train noise level results in a lower SNR. On the other hand, it is clear from Table 3 that the higher overall noise level does not affect response times.

No effect of test condition on task accuracy was found. More in particular, the correct answer rates for the calculation tasks were 0.89 ± 0.12 , 0.88 ± 0.11 , 0.88 ± 0.12 and 0.89 ± 0.12 , respectively for the CBaseline, CTrain, CDutch and CChinese conditions. A paired t -test showed no significant increase or decrease comparing the CTrain, CDutch and CChinese conditions to the CBaseline condition ($\alpha = 0.05$). The correct answer rates for the recall part of the memory task (0.52 ± 0.08 , 0.52 ± 0.08 , 0.51 ± 0.07 and 0.53 ± 0.09) also did not show significant differences. Overall, it can be concluded that, in the present experiment, a change in sound environment influenced task duration rather than task accuracy, contrasting earlier findings [42].

3.4. *Relation between acoustical parameters and task performance*

In the previous section, the effect of overall stimuli properties was considered. Here, we will look at specific acoustical parameters of the stimuli. As stated in Section 3.3, the train noise might mask the interfering speech sound and hence

Table 3: Duration within the calculation task, for 50% and 100% completion of the task, where ”*” indicates statistical significant difference compared to the sample without speech in the same group of 3 stimuli ($p < 0.05$). Standard deviations are shown inside parentheses.

Condition	Dutch-speaking groups		Chinese-speaking groups	
	50%	100%	50%	100%
S250train	64.6 (6.3)	127.1 (12.1)	67.4 (8.4)	132.1 (13.3)
S250farDut	59.6 (7.6)	120.6 (12.8)	59.0 (6.8)*	117.0 (8.7)*
S250farChn	62.6 (4.0)	122.1 (9.9)	62.8 (5.5)*	121.3 (6.4)*
S250train	58.0 (6.8)	117.1 (13.0)	59.1 (7.0)	119.7 (11.6)
S250nearDut	63.9 (6.6)*	125.5 (13.1)*	62.0 (7.8)	123.4 (11.6)
S250nearChn	54.6 (8.1)	110.9 (14.9)	65.6 (11.3)*	129.7 (20.7)*
S330train	60.4 (6.4)	118.1 (12.2)	59.4 (7.7)	117.4 (12.0)
S330nearDut	58.5 (5.8)	116.4 (11.1)	57.5 (5.8)	116.0 (10.0)
S330nearChn	60.8 (8.1)	119.8 (14.0)	64.7 (6.2)	123.2 (9.5)

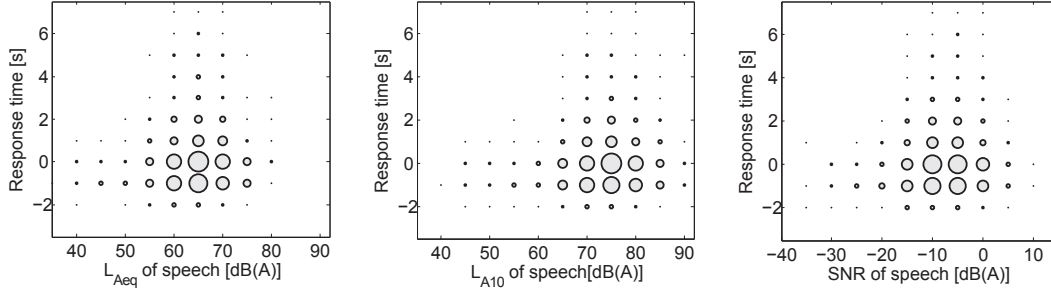


Figure 5: Relation between acoustical parameters and response time. The diameter stands for the number of clicks in a 5dB interval.

have a beneficial effect on task performance. In order to investigate the effect of the acoustical parameters of the background noise on task performance in more detail, the L_{Aeq} , L_{A10} and signal-to-noise ratio of the speech are calculated for each time segment between two subsequent responses. Figure 5 shows the relationship between these acoustical parameters and response time (relative to the baseline for each participant) for the calculation task. It can be seen that the selected acoustical parameters can only partially explain the observed variance in response time; no significant (linear) relationship can be discerned. Consequently, a more in-depth analysis of the background noise is needed.

The influence of background noise on task performance could be explained by how much the noise will distract attention away from the task at hand. The effects not only relate to the noise level, but also to other sound features. In situations where the sound level is rather low, particular spectro-temporal changes in the background sound still may make it salient and thus may make it attract attention. On the other hand, people may adapt to constant broadband noise, even if it has a relatively high level. Therefore, a more sophisticated notice-event model [35, 47] is used to estimate the amount of auditory attention that is paid to the background

noise, and thus not spent on the calculation task at hand. In the following, a short functional description of the model will be given; a more complete mathematical description of the model is beyond the scope of this paper, but can be found in [47]. The model simulates how listeners switch their attention over time between the task and the different sounds present in the acoustic environment, based on bottom-up and top-down cues. The bottom-up cues are determined by the time-dependent saliency of each sound. The latter are calculated on the basis of an auditory saliency map, which accounts for the audibility of each sound over time (it could be temporarily masked by the other sounds) as well as the spectro-temporal structure of the sound (for example, sudden events potentially attract attention). The bottom-up cues are thus purely acoustical, i.e. the language and meaning of speech is not taken into account. Top-down cues are determined by the amount of volitional focusing on particular sounds and on the task at hand, and account for habituation to sounds. A competitive winner-takes-all mechanism finally balances bottom-up and top-down attention, and delimits, on a statistical basis, the time periods during which particular sounds are paid attention to.

For each participant and for each part of the calculation task in which a stimulus combining train noise with speech was played back (i.e. the CDutch and CChinese conditions), the time periods during which the sound of the stimulus would most likely have attracted attention are estimated using the described notice-event model. In particular, the model takes the sound of the stimulus (both the train noise and the speech sound) as input, and gives the time periods that attention is potentially paid to the background sound as output. Subsequently, at four time instances within the test (after 60 s, 100 s, 140 s and 180 s), the total duration that auditory attention was potentially spent to the background sound up to then is

calculated, and compared to the calculation speed. This speed is measured as the number of elementary sums that were performed divided by the given time frame.

Figure 6 shows the calculation speed within the calculation task as a function of the average estimated duration of auditory attention attracted by the background sound. Results are averaged over all participants within each language group, and over all stimuli within the CDutch and CChinese conditions. No significant difference between the two language groups was found. Two main conclusions can be drawn. First, at each of the four time instances within the task, the same trend is visible: the more attention is paid to the noise as predicted by the model, the lower the normalized progress is, which means the participants need more time to respond. This indicates that the attention estimated using the notice-event model could be a good predictor for response times. Second, with increasing time within the task, the slope of the regressed line decreases. This could indicate that the participants were gradually becoming used to the background noise.

3.5. *Analysis of annoyance*

Persistent exposure to noise and the related task interference might have an adverse impact on perceived well-being. This potential effect is analyzed using the answers on the emotion scale (see Figure 4) asked during the experiment; results are shown in Figure 7. The reference bars denote the emotion state before the start of the experiment, while the other two bars indicate the perceived emotion state after each part of the respective tasks. A *t*-test reveals that participants are significantly less pleased after exposure to either of the stimuli of the CTrain, CDutch or CChinese conditions ($p < 0.001$), compared to baseline condition, but no significant differences were found between any of the combinations, CTrain and CDutch/CChinese, with $p > 0.1$. The question asked to reveal the current

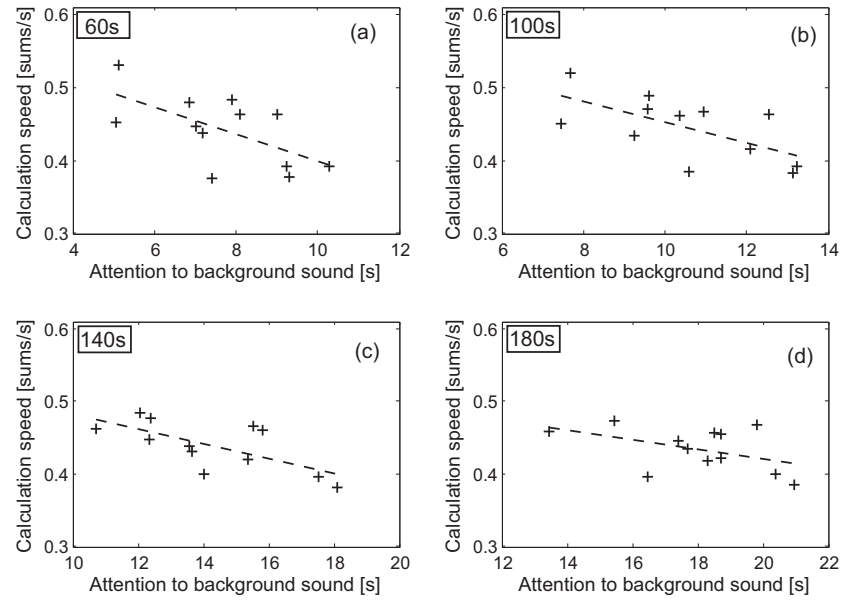


Figure 6: Calculation speed as a function of the average estimated duration of auditory attention spent to the background sound, at four time instances within the task. Each marker represents a combination of language group (Dutch or Chinese mother tongue) and stimulus within the CDutch and CChinese conditions.

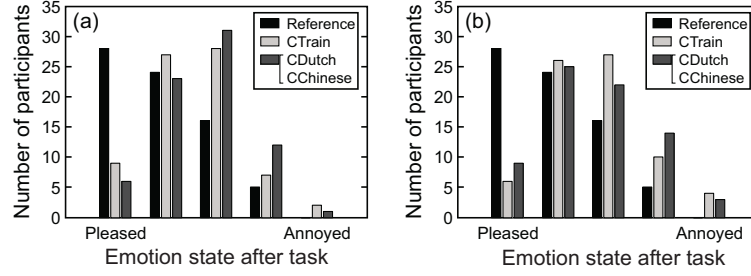


Figure 7: Perceived emotion state after completion of (a) each part of the calculation task, and (b) each recall part of the memory task, under different background noise conditions. The reference bars denote the level of well-being before the start of the particular task.

feeling of the participant did not mention noise, however participants may have guessed that the noise was an important part of the experiment.

Previous research has shown that changes of sound characteristics, such as pitch, timbre and tempo, are more important than changes of sound level in disrupting participants [48, 24]. However, the sound level of the train noise was about 10 dB(A) higher than the sound level of the speech, and therefore was expected to be the main factor dominating perceived annoyance. To confirm the role of sound level and characteristics, an additional questionnaire was administered at the end of the experiment. At the end of the test, the participants were asked “Which noise made you feel more annoying? The train noise, the speech noise, or were these nearly equally annoying?”. Most participants (71 %) felt that the train noise and the speech were about equally annoying. Nevertheless, as much as 22 % found the speech noise to be more annoying, and only 7 % found the train noise to be more annoying. This result somewhat contrasts the findings obtained through the emotional manikins, and indicates that, even under a background noise with a

considerably higher sound level, interfering speech can still lead to a higher annoyance. Consequently, only lowering the sound level of the train noise may not be beneficial as people might also be disturbed by the interfering speech.

4. Discussion and Conclusions

This paper reported on a laboratory experiment in which the influence of background noise inside high-speed trains on task performance and subjective well-being has been investigated. Task accuracy and response times were collected on a series of calculation and verbal memory tasks, performed by participants while they were exposed to train compartment noise stimuli. The latter were constructed through both measurements and simulations, and consisted of a combination of train noise and speech sound.

It was found that response time and accuracy can be considered as independent variables for both types of tasks. The strongest effects of the background noise were found on response times. More in particular, effects were found to be mainly due to the distracting effect of the speech, with no significant influence of overall sound level. The largest interference effect was found when the language of the presented speech was the participants mother tongue, but no significant differences were observed between the participants of each language group. In order to investigate the effect of attention distraction more in detail, a model of saliency-based auditory attention was used to estimate the potential distraction by noise, based on the acoustic stimuli. It was found that response times were positively correlated with the estimated amount of attention that was paid to the background sound. The attention estimated with the model (which accounts for the saliency of the sound and for temporal effects such as habituation) is a bet-

ter predictor for response time than basic sound level indicators. It was found that participants were gradually becoming used to the background noise over the course of the task. All the background noises made participants less pleased and more annoyed, but no significant dependence on the composition of the noise was found. When asked about the annoying effect of the background sound after completion of the whole experiment, with explicit reference to the presence of speech, the speech was sometimes found to be more disturbing than the train noise, but very rarely less disturbing. As participants were strongly distracted from listening to the speech during this experiment and as speech levels were very low compared to the train noise, this experiment shows that sound characteristics and meaning play an important role in perceived annoyance.

The test design that was used showed two main limitations. First, the assessment of emotional state or well being during the test was not very subtle. Although it revealed some effect of the experiment as a whole, subtle differences between different types of noise may have been missed. Yet, it is difficult to question participants about emotion and perception without putting the focus of the participants on the noise rather than on the task at hand. Second, because the recall parts of the memory task involved the participants having to type the answers (similar to [27], where participants answered in writing), response times were considerably influenced by the the persons ability to type and the length of the answer. Consequently, a detailed analysis of the results of the recall parts was not included in this work. Future experiments could consider to use a recall method that is less prone to time, such as using multiple choices for the answers on the questions, or by answering aloud. Furthermore, the influence of the visual setting, which could have a mediating effect on the emergence of annoyance [49], was neglected in the

present experiment.

5. Acknowledgment

Weigang Wei is a beneficiary of a China Scholarship Council (CSC) grant; the support of the CSC is gratefully acknowledged. Both Annelies Bockstael and Bert De Coensel are postdoctoral fellows of the Research Foundation–Flanders (FWO–Vlaanderen); the support of this organisation is gratefully acknowledged as well. The authors would also like to thank Prof. Catherine Lavandier for her beneficial suggestions regarding the experimental design.

6. References

- [1] R. A. Smith, The Japanese Shinkansen: Catalyst for the renaissance of rail, *J. Transp. Hist.* 24 (2003) 222–237.
- [2] R. Vickerman, High-speed rail in Europe: Experience and issues for future development, *Ann. Reg. Sci.* 31 (1997) 21–38.
- [3] F. T. Najafi, F. E. Nassar, Comparison of high-speed rail and maglev systems, *J. Transp. Eng.* 122 (1996) 276–281.
- [4] Y. Moritoh, Y. Zenda, K. Nagakura, Noise control of high speed shinkansen, *J. Sound Vib.* 193 (1996) 319–334.
- [5] T. Kitagawa, K. Nagakura, Aerodynamic noise generated by shinkansen cars, *J. Sound Vib.* 231 (2000) 913–924.
- [6] K. Nagakura, Localization of aerodynamic noise sources of Shinkansen trains, *J. Sound Vib.* 293 (2006) 547–556.

- [7] V. V. Krylov (Ed.), Noise and vibration from high-speed trains, Thomas Telford Publishing, London, UK, 2001.
- [8] C. Talotte, P. E. Gautier, D. J. Thompson, C. Hanson, Identification, modelling and reduction potential of railway noise sources: a critical survey, *J. Sound Vib.* 267 (2003) 447–468.
- [9] C. Mellet, F. Létourneaux, F. Poisson, C. Talotte, High speed train noise emission: Latest investigation of the aerodynamic/rolling noise contribution, *J. Sound Vib.* 293 (2006) 535–546.
- [10] P.-E. Gautier, F. Poisson, F. Létourneaux, High speed trains external noise: recent results in the TGV case, in: *Proc. of the 19th International Congress on Acoustics (ICA)*, Madrid, Spain.
- [11] J. Lambert, P. Champelovier, I. Vernet, Annoyance from high speed train noise: A social survey, *J. Sound Vib.* 193 (1996) 21–28.
- [12] B. van Wee, R. van den Brink, H. Nijland, Environmental impacts of high-speed rail links in cost-benefit analyses: a case study of the Dutch Zuider Zee line, *Transp. Res. Part D* 8 (2003) 299–314.
- [13] B. De Coensel, D. Botteldooren, B. Berglund, M. E. Nilsson, T. De Muer, P. Lercher, Experimental investigation of noise annoyance caused by high-speed trains, *Acta Acust. Acust.* 93 (2007) 589–601.
- [14] E. Parizet, N. Hamzaoui, J. Jacquemoud, Noise assessment in a high-speed train, *Appl. Acoust.* 63 (2002) 1109–1124.

- [15] G. Xiaoan, Railway environmental noise control in china, *J. Sound Vib.* 293 (2006) 1078–1085.
- [16] L. Ding, VIP seats scale analysis and modeling human-machine design study, Master's thesis, Southwest Jiaotong University, Department of Industrial Design, 2011.
- [17] X. Chen, Research of comprehensive evaluation model on riding comfort of high-speed train, Ph.D. thesis, Southwest Jiaotong University, Department of Industrial Design, 2010.
- [18] A. Haapakangas, E. Kankkunen, V. Hongisto, P. Virjonen, D. Oliva, E. Keskinen, Effects of Five Speech Masking Sounds on Performance and Acoustic Satisfaction. Implications for Open-Plan Offices, *Acta Acust. Acust.* 97 (2111) 641–655.
- [19] V. Hongisto, A model predicting the effect of speech of varying intelligibility on work performance, *Indoor Air* 15 (2005) 458–468.
- [20] S. Kuwano, S. Namba, T. Okamoto, Psychological evaluation of sound environment in a compartment of a high-speed train, *J. Sound Vib.* 277 (2004) 491–500.
- [21] K. Zimmer, J. Ghani, W. Ellermeier, The role of task interference and exposure duration in judging noise annoyance, *J. Sound Vib.* 311 (2008) 1039–1051.
- [22] J. Bengtsson, K. Persson Waye, A. Kjellberg, Evaluations of effects due to low-frequency noise in a low demanding work situation, *J. Sound Vib.* 278 (2004) 83–99.

- [23] J. S. Little, F. H. Martin, R. H. S. Thomson, Speech versus non-speech as irrelevant sound: Controlling acoustic variation, *Biol. Psychol* 85 (2010) 62–70.
- [24] D. Jones, C. Madden, C. Miles, Privileged access by irrelevant speech to short-term memory: The role of changing state, *Q. J. Exp. Psychol.* 44A (1992) 645–669.
- [25] W. J. Macken, F. G. Phelps, D. M. Jones, What causes auditory distraction?, *Psychon. Bull. Rev.* 16 (2009) 139–144.
- [26] W. Ellermeier, J. Hellbrück, Is level irrelevant in "irrelevant speech"? effects of loudness, signal-to-noise ratio, and binaural unmasking, *J. Exp. Psychol.-Hum. Percept. Perform.* 24 (1998) 1406–1414.
- [27] C. J. P. Oswald, S. Tremblay, D. M. Jones, Disruption of comprehension by the meaning of irrelevant sound, *Memory* 8 (2000) 345–350.
- [28] C. P. Beaman, N. J. Holt, Reverberant auditory environments: The effects of multiple echoes on distraction by 'irrelevant' speech, *Appl. Cogn. Psychol.* 21 (2007) 1077–1090.
- [29] S. Tremblay, A. P. Nicholls, D. Alford, D. M. Jones, The irrelevant sound effect: Does speech play a special role?, *J. Exp. Psychol.-Learn. Mem. Cogn.* 26 (2000) 1750–1754.
- [30] N. Venetjoki, A. Kaarlela-Tuomaala, E. Keskinen, V. Hongisto, The effect of speech and speech intelligibility on task performance, *Ergonomics* 49 (2006) 1068–1091.

- [31] E. M. Elliott, K. M. Barrilleaux, N. Cowan, Individual differences in the ability to avoid distracting sounds, *Eur. J. Cognit. Psychol.* 18 (2006) 90–108.
- [32] W. Ellermeier, K. Zimmer, Individual differences in susceptibility to the ”irrelevant speech effect”, *J. Acoust. Soc. Am.* 102 (1997) 2191–2199.
- [33] W. Ellermeier, M. Eigenstetter, K. Zimmer, Psychoacoustic correlates of individual noise sensitivity, *J. Acoust. Soc. Am.* 109 (2001) 1464–1473.
- [34] E. Öhrström, M. Björkman, R. Rylander, Noise annoyance with regard to neurophysiological sensitivity, subjective noise sensitivity and personality variables, *Psychol. Med.* 18 (1988) 605–613.
- [35] B. De Coensel, D. Botteldooren, T. De Muer, B. Berglund, M. E. Nilsson, P. Lercher, A model for the perception of environmental sound based on notice-events, *J. Acoust. Soc. Am.* 126 (2009) 656–665.
- [36] M. Sneddon, K. Pearsons, S. Fidell, Laboratory study of the noticeability and annoyance of low signal-to-noise ratio sounds, *Noise Control Eng. J.* 51 (2003) 300–305.
- [37] T. Sato, T. Yano, M. Bjorkman, R. Rylander, Road traffic noise annoyance in relation to average noise level, number of events and maximum noise level, *J. Sound Vib.* 223 (1999) 775–784.
- [38] J. D. Bransford, J. J. Franks, The abstraction of linguistic ideas: A review, *Cognition* 1 (1972) 211–249.

- [39] S. Banbury, D. C. Berry, Disruption of office-related tasks by speech and office noise, *Br. J. Psychol.* 89 (1998) 499–517.
- [40] C. P. Beaman, D. M. Jones, Irrelevant sound disrupts order information in free as in serial recall, *Q. J. Exp. Psychol.* 51A (1998) 615–636.
- [41] S. Banbury, D. C. Berry, Habituation and dishabituation to speech and office noise, *J. Exp. Psychol.-Appl.* 3 (1997) 181–195.
- [42] J. L. Szalma, P. A. Hancock, Noise effects on human performance: A meta-analytic synthesis, *Psychol. Bull.* 137 (2011) 682–707.
- [43] C. L. Christensen, Odeon room acoustics program, User manual 10.1, Odeon A/S, Lyngby, Denmark, 2009.
- [44] P. Corthals, Sound pressure level of running speech: percentile level statistics and equivalent continuous sound level, *Folia Phoniatri. Logop.* 56 (2004) 170–181.
- [45] M. M. Bradley, P. J. Lang, Measuring emotion: The self-assessment manikin and the semantic differential, *J. Behav. Ther. & Exp. Psychiat.* 25 (1994) 49–59.
- [46] M. E. Nilsson, B. Berglund, O. Axelsson, Soundscapes perceived by children and adults, in: *Proc. of Fechner Day 2003*, Larnaca, Cyprus, pp. 199–204.
- [47] B. De Coensel, D. Botteldooren, A model of saliency-based auditory attention to environmental sound, in: *Proc. of the 20th International Congress on Acoustics (ICA)*, Sydney, Australia.

- [48] S. P. Banbury, W. J. Macken, S. Tremblay, D. M. Jones, Auditory distraction and short-term memory: Phenomena and practical implications, *Hum. Factors* 43 (2001) 12–29.
- [49] S. Viollon, C. Lavandier, C. Drake, Influence of visual setting on sound ratings in an urban environment, *Appl. Acoust.* 63 (2002) 493–511.